УДК: 543.42+681.3

ZaIR, ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА ПО ИК СПЕКТРОСКОПИИ: БАЗА ДАННЫХ, ПОИСКОВЫЙ АЛГОРИТМ И ОСОБЕННОСТИ ИНТЕРФЕЙСА

И.И.Строков., К.С.Лебедев, М.И.Подгорная, Б.Г.Дерендяев
Новосибирский институт органическойхимии им. Н.Н.Ворожцова СО РАН
630090, Новосибирск, пр.Лаврентьева, 9
der@nioch.nsc.ru

Поступила в редакцию 19 марта 2002 г.

Предложены оригинальные методы поиска и хранения информации в базе данных, содержащей свыше 70000 инфракрасных (ИК) спектров различных химических соединений. Алгоритм поиска обеспечивает быстрый отбор релевантных запросу записей, сравнение полных кривых ИК поглощения или отдельных участков спектра.

Строков Игорь Иванович – научный сотрудник лаборатории программного обеспечения диалоговых систем в химии Новосибирского института органической химии им. Н.Н.Ворожцова Сибирского отделения РАН (НИОХ СО РАН), кандидат химических наук.
Область научных интересов: компьютерные технологии, теория графов, методы установления строения органических соединений по спектральным данным, информационные технологии в химии.
Автор 30 печатных работ.

Лебедев Константин Сергеевич – профессор кафедры аналитической химии Новомосковского института РХТУ им. Д.И. Менделеева, доктор химических наук (в период выполнения данной работы – сотрудник НИОХ СО РАН).
Область научных интересов: компьютерные методы установления строения органических соединений по спектральным данным (МС, ИК, ЯМР), поисковые и экспертные системы, информационные технологии в химии.
Автор 80 печатных работ.

Подгорная Маргарита Ивановна – заведующая лабораторией информационного обеспечения поисковых систем в химии НИОХ СО РАН, кандидат химических наук, старший научный сотрудник.
Область научных интересов: молекулярная спектроскопия, базы данных по структурам и спектрам соединений.
Автор более 100 печатных работ.

Дерендяев Борис Григорьевич – руководитель отдела научно-технический центр по химической информатике НИОХ СО РАН, заведующий лабораторией, доктор химических наук, профессор.
Область научных интересов: физическая и органическая химия, аналитическая химия, химическая информатика, молекулярная спектроскопия, компьютерные методы анализа физико-химических данных.
Автор более 170 печатных работ.

Инфракрасная спектроскопия (ИК) является одним из наиболее доступных и широко используемых в аналитической практике методов установления строения химических соединений. Вместе с этим анализ ИК спектров остается достаточно сложным и трудоемким, требующим специалистов высокой квалификации как в области инфракрасной спектроскопии, так и химии. Не удивительно поэтому, что большое значение приобретают компьютерные средства и методы анализа данных ИК спектроскопии, позволяющие существенно облегчить и ускорить решение многих аналитических задач. Среди разработанных к настоящему времени методов (искусственный интеллект, распознавание образов, искусственные нейронные сети [1]) наиболее простым и эффективным является библиотечный поиск [2]. С помощью этого метода можно успешно решать не только одну из наиболее массовых задач — идентификацию ранее описанных веществ, но и получать надежные сведения о строении неизвестных соединений [3–6]. Для этого используются информационно-поисковые системы (ИПС) на основе баз данных ИК спектров. Принцип рабо-

ты ИПС довольно прост и состоит в сравнении спектра изучаемого соединения с эталонными спектрами, хранящимися в базе данных. Однако, в отличие от других видов молекулярной спектроскопии (например, ЯМР и масс-спектрометрии), при разработке ИПС по ИК спектроскопии возникает ряд трудностей. Во-первых, ИК спектр не является дискретной функцией и имеет достаточно сложный характер. Во-вторых, вид спектра одного и того же соединения может зачастую существенно отличаться и зависит от большого числа факторов (температура и чистота образца, условия регистрации, тип прибора и т.п.). В-третьих, базы данных по ИК спектрам занимают очень много дисковой памяти, что сильно сказывается на времени поиска. В связи с этим вопросы компьютерного анализа, сравнения и представления ИК спектров продолжают привлекать внимание разработчиков ИПС [7].

В данной статье описывается разработанная в Новосибирском институте органической химии СО РАН информационно-поисковая система ZaIR на основе базы данных, содержащей более 70000 полных ИК спектров различных химических соединений. При этом основное внимание уделено рассмотрению технических аспектов, во многом определяющих «пользовательские» качества системы, а следовательно, и возможность широкого применения в аналитической практике. Это относится к способам представления и хранения ИК спектров, алгоритму поиска и интерфейсу пользователя.

Метод хранения спектральных кривых

В общем случае полный ИК спектр описывается двумерной кривой, задающей интенсивность поглощения сигнала (ось ординат) от его частоты (ось абсцисс). Чаще всего, например в принятом ИЮПАК формате обмена и хранения JCAMP-DX [8,9], спектр представлен в виде массива ординат точек, взятых через равные интервалы частот поглощения. Поскольку интервал частот, как правило, равен 2÷4 $см^{-1}$, число точек, описывающих спектр в диапазоне 200÷4000 $см^{-1}$, может быть близко к 2000. Если база данных содержит десятки тысяч спектров, то их хранение в данном формате требует больших затрат памяти. Поэтому в крупных базах данных используют различные способы сжатия спектральной информации.

Известно, например, упрощённое кодирование спектра массивом бит (1, если есть поглощение заданного уровня в фиксированном интервале, 0 — если нет). Более распространено представление спектра в виде множества пиков, каждый из которых описывается несколькими параметрами, отражающими положение, интенсивность и полуширину пиков. Поскольку пиков в спектре намного (примерно в 100 раз) меньше, чем точек, описывающих полную кривую поглощения, то такой способ обеспечивает значительную экономию памяти. К сожалению, при этом неизбежно теряются многие тонкие особенности кривой в целом, которые в ряде случаев могут быть полезны при определении строения соединения, оценках его чистоты и качества спектра. Поэтому большинство пользователей ИПС предпочитает работать с БД, содержащими по возможности более полную информацию об исходной кривой поглощения.

Форма кривой сохраняется лучше при её частотном анализе и приближенном представлении в виде суммы периодических или импульсных функций, подобно тому, как это делается при сжатии изображений в технологии JPEG. Однако в этом варианте сжатия информации экономия памяти меньше, чем в случае описания спектра набором пиков. Проблема с применением данного приёма состоит в том, что ИК спектр чаще всего не является «гладким», то есть имеет мало повторений в периодичности или форме пиков. Поэтому, в отличие от фотографий и рисунков, здесь обычно не удаётся совместить большую степень сжатия информации с минимальной потерей качества.

Таким образом, и многолетний опыт использования предыдущих версий ИПС по ИК спектроскопии, и отмеченное выше свидетельствуют, что в случае создания систем на основе больших баз данных оказывается оправданным хранение спектров без потери информации. Рассмотрим, как это достигается при формировании БД системы ZaIR.

Исходная спектральная кривая в данном случае представлена как массив ординат точек, полученный путём сканирования спектра с бумажного носителя и последующего распознавания кривой [10]. Ординаты точек заданы числами от 0 до 255, что лишь немного меньше, чем разрешение сканирующего устройства. Абсциссы точек отстоят друг от друга на 1.6(6) $см^{-1}$ (то есть, 6 точек на 10 $см^{-1}$) в области «отпечатков пальцев» 200–2000 $см^{-1}$ и вдвое больше — 3.3(3) $см^{-1}$ - в области 2000–4000 $см^{-1}$, что соответствует числу точек в спектре 1680 или больше, если общий диапазон спектра превышает указанный. В силу размещения значений абсцисс через равные интервалы, для каждой точки достаточно задать её ординату, которая умещается в 1 байт. Таким образом, исходный спектр — это массив, содержащий ~1700 чисел (байт). Его последующее сжатие проводится в два этапа.

На первом этапе массив данных записывается в виде набора «команд», то есть, каждый байт записи рассматривается как команда. Большинство команд, однако, просто означают число, равное значению самой команды (например, команда 50 означает число 50, 72 — число 72 и т.д.). Некоторые особые команды означают повторение одного или нескольких чисел, ряд возрастающих или убывающих чисел. С помощью таких команд можно компактнее записать многие регулярные участки кривой (если они есть). В среднем использование команд позволяет сжать массив данных почти вдвое, хотя в ряде случаев запись набора команд может оказаться длиннее, чем исходного представления.

На втором этапе выполняется дополнительное сжатие: набор команд записывается более кратко с помощью целых чисел переменной длины. То есть более часто встречающиеся команды записываются целыми числами, которые состоят из меньшего количества бит (длина числа определяется его начальными битами)[1]. Этот способ, который обычно рассматривают как упрощённый метод Хаффмана, сокращает запись ещё примерно на 1/3. В итоге для записи спектра из ~1700 точек (в среднем по базе данных) требуется чуть меньше 600 байт. Важно, что обратная распаковка записи в исходный массив ординат не требует сложных вычислений и происходит очень быстро.

Индексный файл для быстрого поиска

Полные кривые сохраняют максимум исходной информации о спектре, что важно не только для его зрительного восприятия, но позволяет проводить и детальное сопоставление кривых, например, при поиске в базе данных спектра, тождественного заданному в запросе. С другой стороны, такое сравнение (и, следовательно, поиск) трудно сделать быстрым — хотя бы из-за большого числа точек, описывающих кривую. Более того, поисковый запрос в общем случае может включать не полные кривые, а их отдельные участки, пики и полосы спектра. Здесь под пиком понимается участок интенсивного поглощения кривой, выделяющийся на фоне соседних участков (обычно — локальный максимум), а под полосой — широкий пик или несколько слившихся пиков. Очевидно, что поиск в этих случаях можно ускорить, если заранее вычислить положения спектральных пиков и основные параметры их формы (высота, ширина) для всех спектров базы данных. Отметим, однако, что не каждый локальный максимум отвечает пику (шумовой сигнал, состоящий из мелких всплесков) и, наоборот, пик может не проявиться в виде максимума кривой — например, при слиянии с другими пиками. Алгоритм определения пиков спектральной кривой должен по возможности учитывать эти случаи.

В системе ZaIR используется достаточно сложный алгоритм, включающий набор параметров, которые настроены так, чтобы результат выделения пиков в разнообразных ситуациях оказался наилучшим с точки зрения эксперта-спектроскописта.

Пик считается выделенным, если в спектре найдены три точки, отвечающие вершине пика и ближайшим впадинам по обеим сторонам вершины (рис. 1). Далее, однако, для целей поиска пик описывается не этими тремя точками, а четырьмя параметрами, которые отвечают: x — *положению* пика (абсцисса его вершины), h — *высоте* (ордината вершины), u — *возвышению* над уровнем кривой в месте положения пика (средняя разница ординат вершины и впадин), a — *углу*, образуемому вершиной и впадинами. Положение пика выражается в $см^{-1}$, а три остальные параметра — безразмерными величинами (целыми числами от 0 до 3).

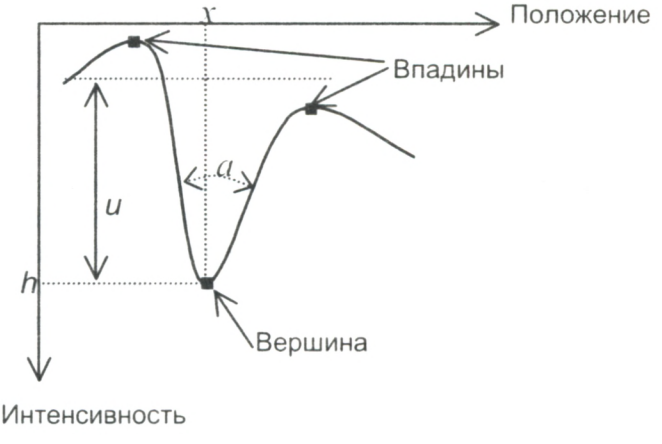

Рис.1. Параметры пика

В то время как положение пика совпадает с абсциссой его вершины, для вычисления трех других параметров используются интервалы исходных значений ординат (для u и h) или углов (для a). В этих случаях полные диапазоны исходных значений от 0 до 255 для ординат и от 0 до 180^0 — для углов разбиваются на четыре интер-

[1] Рассмотрим для примера массив из двадцати 8-битовых чисел: 200, 201, 200, 100, 202, 203, 200, 200, 207, 200, 200, 204, 100, 205, 100, 100, 206, 100, 100, 200. Запишем числа 200 и 100 (они встречаются чаще других, 13 раз) набором битов **10** и **11** соответственно. Оставшиеся числа 201–207 запишем битами как **0000**, **0001** ,..., **0111**. Всего понадобится тринадцать 2-битных чисел и семь 4-битных чисел, что составит в сумме 13·2 + 7·4 = 54 бита против 20·8 = 160 в исходном массиве. Поскольку 2-битные числа начинаются с бита **1**, а 4-битные — с бита **0**, эти числа можно различить и, следовательно, восстановить по ним исходный массив.

вала так, чтобы в каждый из них попало одинаковое число пиков из спектров базы данных. Например, всего во всех спектрах выделено без малого 2 миллиона пиков. Четверть из них (500000) имеют ординату вершин в интервале 0–54 и им соответствует $h = 0$; ординаты вершин ещё 500000 пиков попадают в интервал 55–99, и им присваивается $h = 1$, и т.п. Данные об интервалах исходных значений ординат и углов для всех значений параметров приведены в табл. 1.

Таблица 1

Интервалы исходных значений ординат и углов для всех значений параметров пиков

Значение параметра	Интервалы исходных значений		
	Для h	Для u	Для a
0	0–54	0–15	0–13°
1	55–99	16–38	14–33°
2	100–161	39–82	34–80°
3	162–255	83–255	81–180°

Далее все четыре параметра объединяются в одно число, называемое дескриптором пика d:

$$d = 64x + 16h + 4u + a.$$

Таким образом, дескриптор содержит все параметры пика, причём положение составляет самую старшую часть числа, а угол — самую младшую. Например, параметры и дескрипторы пиков, выделенных из спектра на рис. 2, приведены в табл. 2.

Таблица 2

Параметры и дескрипторы пиков, показанных на рис.2

X	h	u	a	d
386	1	1	3	24727
591	1	1	3	37847
730	1	0	3	46739
920	3	3	1	58941
1086	3	3	2	9566
1219	3	2	2	78074
1364	1	0	3	87315
1438	1	2	2	92058
1627	0	1	3	104135
1780	3	3	1	113981
1861	2	3	1	119149
2102	0	0	3	134531
2839	1	1	2	181718
2942	2	2	2	188330
3416	1	1	3	218647

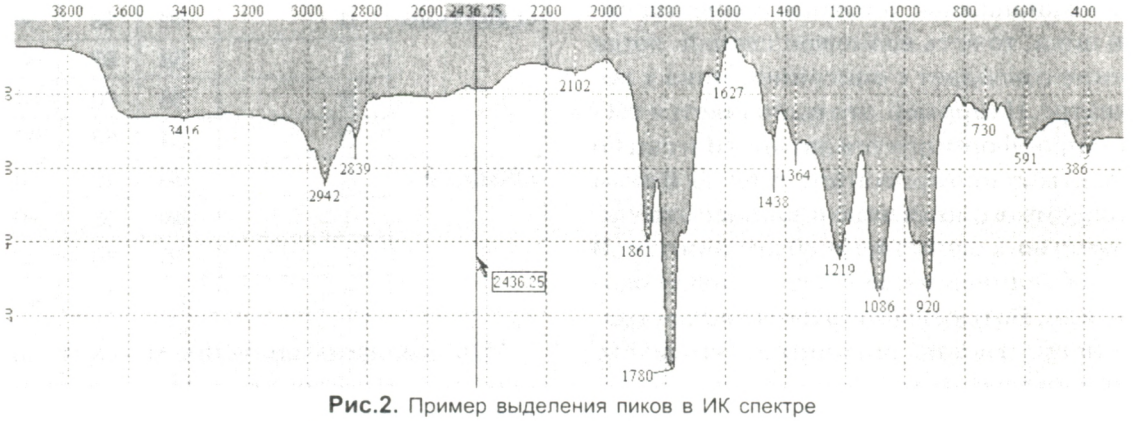

Рис.2. Пример выделения пиков в ИК спектре

Для ускорения спектрального поиска информация о дескрипторах всех пиков, выделенных в спектрах БД, записывается в отдельный индексно-последовательный файл формата JoKey [11], который обеспечивает прямой доступ к записям по ключу. Ключ записи в индексном файле — это дескриптор пика d, а сама запись содержит номера всех спектров, в которых имеется пик с таким же дескриптором.

Рассмотрим, как используется индексный файл для быстрого поиска спектральных данных. Поисковый запрос в общем случае — это набор *ограничений* на отдельные пики. Каждое ограничение включает точные значения или диапазоны всех параметров пика. Например, ограничение [x = 1600–1620, h = 3, u = 3, a = 0–3] отвечает очень высокому пику (h = 3, u = 3) с любым углом (a = 0–3) и положением вершины в шкале частот поглощения от 1600 до 1620 см$^{-1}$. Очевидно, что каждому из ограничений можно сопоставить множество дескрипторов пиков, параметры которых попадают в заданные диапазоны. Каждому дескриптору, в свою очередь, отвечает множество спектров, которые находятся прямым доступом в индексном файле. Объединение таких множеств образует множество спектров, удовлетворяющих данному ограничению. Далее, если объединённые множества, найденные для всех ограничений зап-

роса, пересечь, то получится множество спектров, удовлетворяющих запросу в целом (рис.3).

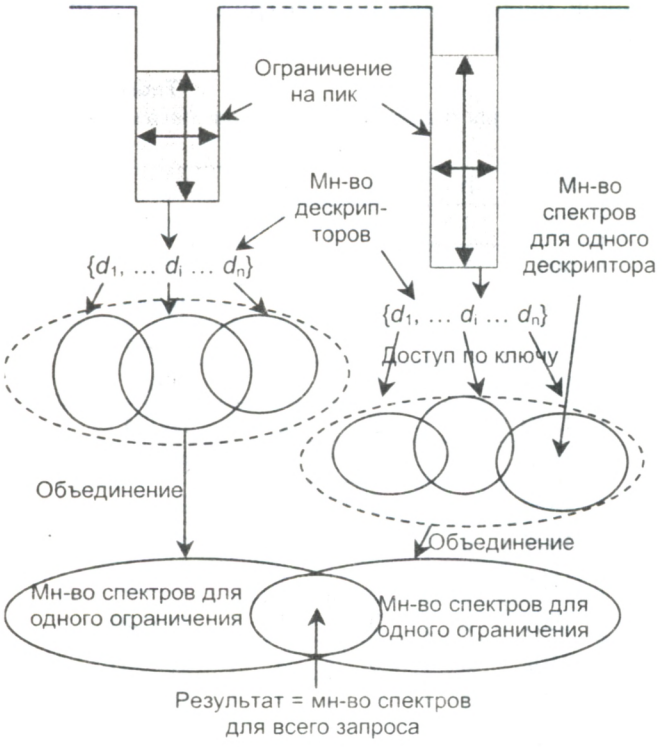

Рис.3. Общая схема спектрального поиска (примечание: мн-во – множество)

Заметим, что ограничения в запросе могут включать не диапазоны, а точные значения параметров пика, то есть возможен случай, когда ограничение совпадает с описанием пика в искомом спектре. Например, это выполняется тогда, когда запрос формируется автоматически по полной спектральной кривой (см. табл. 2). В этом случае каждому ограничению в запросе отвечает не множество, а единственный дескриптор. В результате объединенное множество может оказаться пустым. Во избежание этого каждое ограничение искусственно расширяется: вместо единственного дескриптора (отвечающего выделенному пику) в качестве ключей предъявляется множество близких по значениям дескрипторов. Формально это означает, что положение пика в отобранном спектре может отличаться от положения пика в спектре запроса (далее, пик запроса) не более чем на заданную величину — примерно на 20 $см^{-1}$, а высота, возвышение и угол могут принимать любые из возможных четырех значений. Иными словами, отбор идёт не на точное, а приблизительное совпадение дескрипторов пиков. Чтобы в этих условиях соблюсти преимущество пиков, более близких пику запроса, разница соответствующих дескрипторов (обозначим их d_i и d_j) учитывается с помощью оценки $p_{ij} = f(d_i, d_j)$.

Таким образом, для каждого спектра базы данных в процессе поиска накапливается оценка p (называемая также фактором совпадения), согласно которой он занимает место в списке результатов поиска. Оценка p выражается формулой $p = \sum p_{ij}$, где каждое слагаемое p_{ij} отвечает i-му пику запроса и j-му пику спектра из базы данных, дающему максимальное p_{ij} для данного i То есть для каждого пика запроса выбирается только один, наилучшим образом совпадающий с ним пик в спектре базы данных.

Уточним теперь, как рассчитываются оценки p_{ij}. В предположении о независимом влиянии на оценку параметров пиков, p_{ij} выражается произведением: $p_{ij} = p_x \cdot p_h \cdot p_r \cdot p_a$, где p_x отвечает за разницу положений пиков, p_h — высоты и т.д. Каждый из множителей p_x, …., p_a не вычисляется, а извлекается из таблиц заранее заданных значений, где колонки таблицы отвечают значению параметра в пике запроса, а строки — в пике спектра базы данных. Для примера ниже приводится табл.3 для сравнения высот пиков h в разных режимах поиска.

Таблица 3

Таблицы сравнения высот пиков h в разных режимах поиска

Режим	Высота пика	Высота пика запроса			
	спектра	0	1	2	3
«Обычный»	0	60	58	56	54
	1	58	62	60	58
	2	56	60	64	62
	3	54	58	62	64
«Жесткий»	0	50	40	10	5
	1	40	55	40	10
	2	10	40	60	40
	3	5	10	40	64

Максимальное значение множителя в этой и в других таблицах для ускорения вычислений выбрано 64 (2^6), а минимальное зависит от режима поиска. Исходя из значения 64, можно найти максимальную оценку при сравнении двух пиков (64·64·64·64 = 16777216) и, далее, двух спектров, что позволяет нормировать итоговую оценку отбираемого из базы данных спектра. Таблицы множителей позволяют обойтись без аналитических выражений и, таким образом, ускорить вычисления. Кроме того, они обеспечивают желаемую гибкость при настройке поисковых параметров, но, с другой стороны, затрудняют их оптимизацию. В то же время процедура оптимизации выполняется один раз разработчиками системы, а пользователю предоставляются три основных режима поиска (табл.4).

Таблица 4

Основные режимы поиска

Режим	Характер задачи
«Обычный»	Идентификация индивидуального соединения по ИК спектру
«Жесткий»	Детальное сравнение спектральных кривых или отдельных ее участков с использованием Евклидовой метрики
«Мягкий»	Анализ спектров смесей соединений («обратный поиск»)

Особенности пользовательского интерфейса

Последовательность действий при работе с системой ZaIR традиционная и включает задание запроса, проведение поиска и просмотр результатов. Описанные выше технические приёмы нацелены на ускорение процедуры поиска для разнообразных поисковых заданий. Сравнительные данные о скорости поиска в двух основных режимах разных поисковых систем приведены в табл.5.

Просмотр результатов поиска, конечно, более простая задача, но в ней тоже есть свои особенности. Прежде всего, результаты поиска — это спектры, которые хранятся в базе данных или используются в качестве запроса, только с добавлением фактора совпадения и возможности графического сравнения с поисковым запросом. Достаточно логично, поэтому, при просмотре результатов использовать тот же интерфейс, что и при задании запроса и листании записей базы данных. Кроме того, такой подход упрощает повторный поиск с использованием результатов предыдущего поиска, например, при анализе спектров смесей веществ.

Таблица 5

Времена поиска по спектру (в секундах) в базе данных из 50000 записей на компьютере Pentium-233 с 64 Мб оперативной памяти

Режим поиска	OMNIC [13]	ИК-ЭКСПЕРТ [12]	ZaIR
«Обычный» (поиск по пикам)	Нет данных	60	0,3
«Жесткий» (Евклидова метрика)	27	123	0,5

Так, например, на рис.4 приведён типичный вид экрана при просмотре результатов поиска. Отличие от других видов экрана сводится к тому, что в окне добавлены спектр запроса, название запроса и фактор совпадения приведённых спектров. Все средства для просмотра данных, листания списка соединений, проведения поиска и т.д. остаются неизменными. Переход же между видами страниц осуществляется переключением «закладок» на левом поле окна.

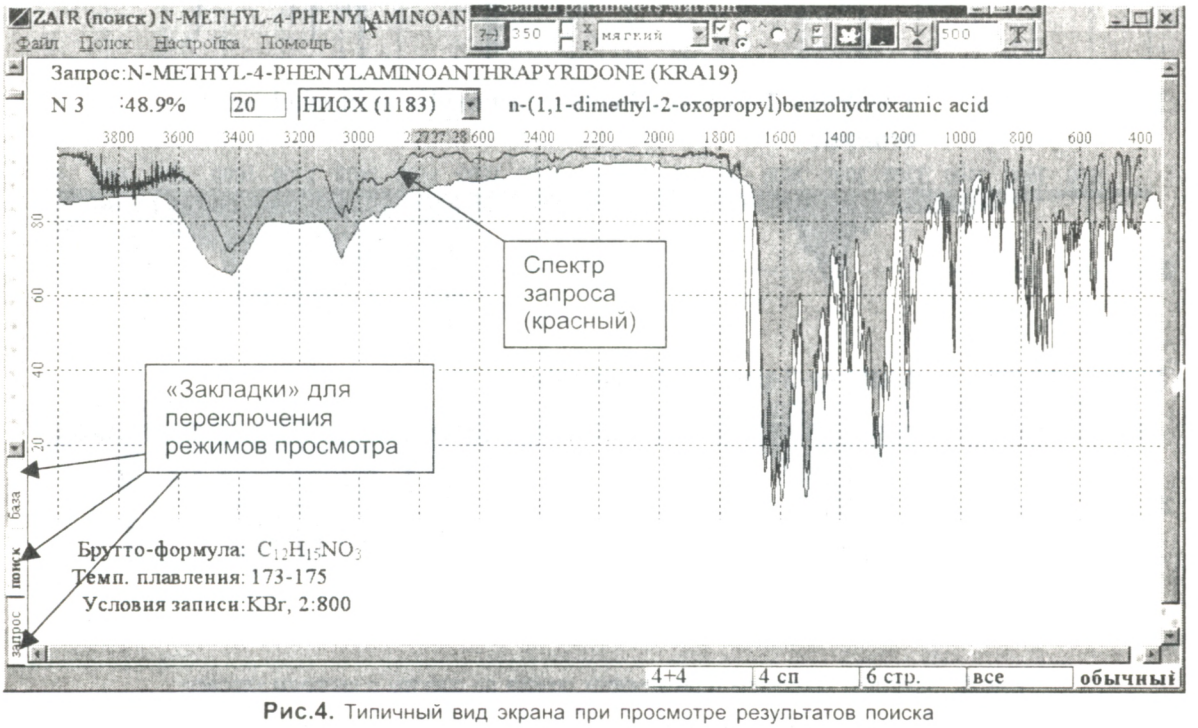

Рис.4. Типичный вид экрана при просмотре результатов поиска

Кроме спектральной кривой, в базе данных для каждого соединения хранятся следующие доступные из литературных источников сведения: химическое название, молекулярный вес, молекулярная формула, структурная формула, физико-химические свойства (температура кипения, плавления и т.п.) и условия регистрации спектров.

В зависимости от содержимого каждой конк-

ретной записи базы данных, результатов поиска, правил составления отчетов и т.д., могут потребоваться разные формы представления данных на экране или на бумаге. Таким образом, возникает вопрос: как управлять расположением графических и текстовых данных на экране без усложнения пользовательского интерфейса и без лишнего программирования?

В системе ZaIR использованы так называемые «виды страниц», которые представляют собой точное описание того, как разместить текстовые поля, прямоугольные окна и другие графические элементы на странице (на экране или бумаге). В отличие от статического размещения в формах (принятых, например, в системах программирования Delphi, Visual Basic и др.), вид страницы зависит от её размера, взаимного расположения, величины и наличия размещаемых элементов, что обеспечивает более гибкую и рациональную компоновку страниц. Можно сказать, что виды страниц содержат элементы типографского набора. В частности, эта особенность используется для экспорта страниц экрана в другие программы работы с документами, например в MS Word.

В качестве иллюстрации на рис.5 приведен фрагмент протокола о результатах поиска по спектру «неизвестного» соединения (шифр «cocaine»). Можно видеть, что первое место в поисковом ответе занимает cocaine hydrochloride, спектр которого довольно хорошо (56,2 %) совпадает со спектром запроса.

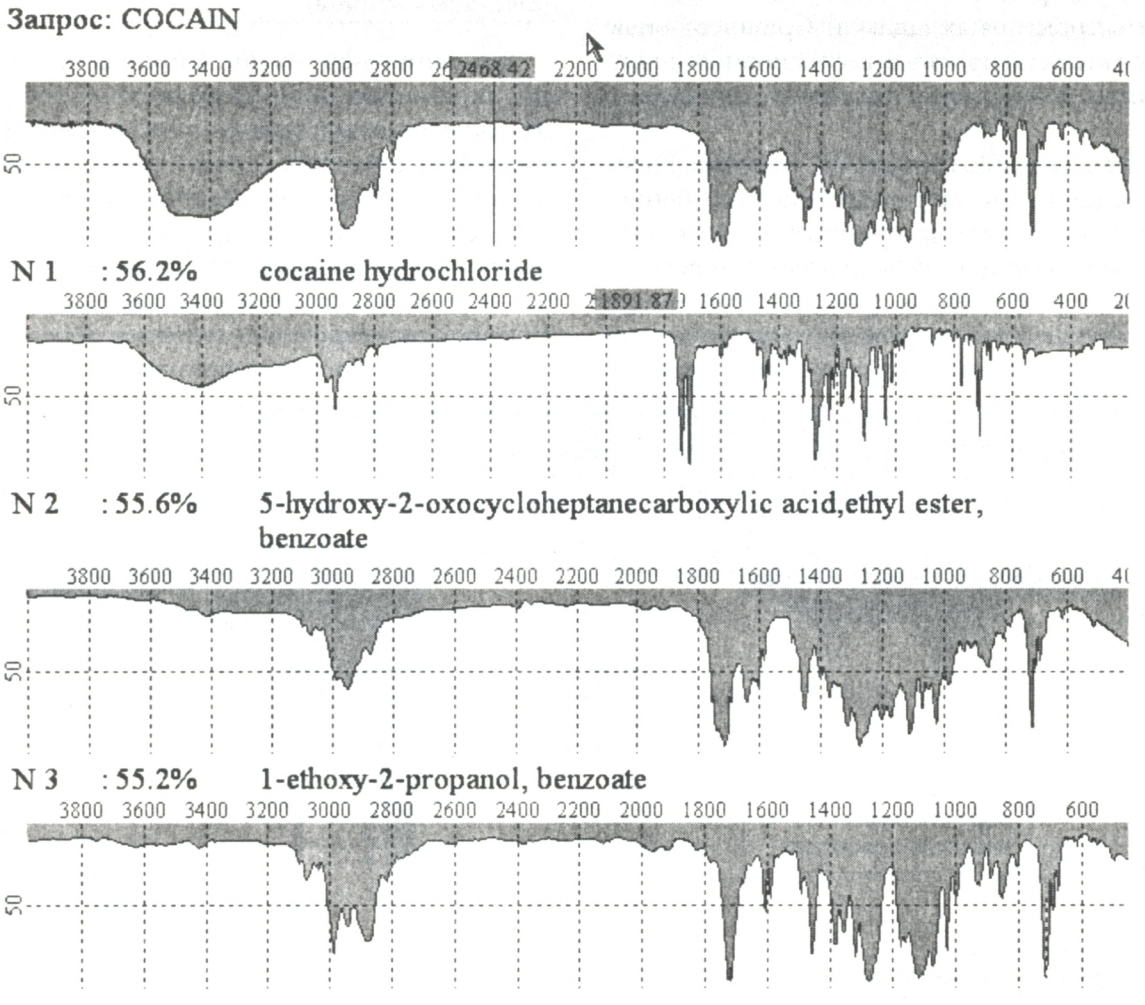

Рис.5. Фрагмент протокола о результатах поиска по спектру с шифром «cocain»

Заметим, что информация на экране может использоваться как для её просмотра, так и для задания поискового запроса. Например, для поиска по названию соединения достаточно ввести в соответствующее поле фрагмент искомого названия — в данном случае изменение поля является признаком того, что оно используется как запрос. В отсутствие изменённых полей запрос составляется на основе пиков, автоматически выделяемых в текущем (то есть присутствую-

щем на экране) спектре. Поиск по текстовым и числовым полям не обсуждается, поскольку основывается на хорошо известных алгоритмах.

В заключение отметим, что система ZaIR имеет структурированную по областям применения и свойствам химических веществ базу данных, сформированную на основе известных зарубежных и отечественных коллекций ИК спектров. В настоящее время база включает 25 разделов (полимеры, лекарства, красители, пластификаторы, пищевые добавки, ПАВ, нефтепродукты, минералы и т.п.) и содержит в общей сложности 74670 ИК спектров.

Реализованный в системе набор программных средств позволяет выполнять следующие основные операции:

- просмотр содержимого отдельных разделов базы данных по номерам;
- чтение спектра запроса в формате JCAMP-DX;
- поиск соединений, спектры которых наиболее близки запросу в режимах «прямого», «обратного» и «комбинированного» методов сравнения спектров;
- поиск по отдельным участкам спектра и полосам поглощения;
- поиск соединений по физико-химическим признакам (название, молекулярный вес, брутто-формула, температура кипения/плавления, условия регистрации спектра и т.п.);
- совместный поиск по спектральным и физико-химическим признакам;
- просмотр и печать результатов поиска;
- экспорт результатов поиска в формат RTF.

Опытная эксплуатация показала, что наиболее эффективно использование системы при решении задач идентификации ранее описанных веществ. При этом положительный результат достигается в различных ситуациях, связанных с чистотой анализируемого образца и условиями регистрации ИК спектров. В то же время, система может оказывать помощь и при установлении строения соединений, спектры которых отсутствуют в базе данных системы. Это достигается путем поиска соединений, спектры которых совпадают со спектром изучаемого соединения по отдельным полосам или имеют общий характер кривых поглощения на отдельных участках спектра. Последующий анализ отобранных из базы данных соединений позволяет сделать правильные суждения о классе изучаемого соединения и определить наличие в его структуре тех или иных фрагментов.

Система ZaIR работает в среде Windows, требует 75 Мбайт дисковой памяти и доступна широкому кругу пользователей (более подробную информацию о системе и вопросам приобретения можно получить по адресу: der@nioch.nsc.ru).

Авторы выражают благодарность Российскому фонду фундаментальных исследований (грант 01-03-32357) за частичную финансовую поддержку данной работы.

ЛИТЕРАТУРА

1. Эляшберг М.Е.//Успехи химии. 1999. Т.68. С.579–604.
2. Warr W.A. // Anal. Chem. 1993. V.65. P.1087A–1095A.
3. Лебедев К.С., Шарапова О.Н., Коробейничева И.К., Кохов В.А. // Изв. СО РАН (Сиб. хим. журн.). 1993. №1. С.50–56.
4. Varmuza K., Penchev P.N., Scsibrany H. // J. Chem. Inf. Comput. Sci. 1998. V.38. P.420–427.
5. Piottukh-Peletsky V. N., Korobeinicheva I. K., Bogdanova T.F., Molodtsov S.G., Derendyaev B.G. // Anal. Chim. Acta. 2000. V.409. P.181–195.
6. Дерендяев Б.Г., Лебедев К.С., Строков И.И., Пиоттух-Пелецкий В.Н., Молодцов С.Г., Подгорная М.И.//Химия в интер.устойчивого развития. 1998. Т.6. С.25–39.
7. Lau O-W., Hon P-K., Bai T. // Vibrational Spectroscopy. 2000. V.23. P.23–30.
8. McDonald R. S., Wilks P. A. // Appl. Spectrosc. 1988. V.3, №2. P.16–18.
9. Gasteiger J., Hendriks B.M.P., Hoever P., Jochum C., Somberg H. // Appl. Spectrosc. 1991. V.45. P.4–11.
10. Подгорная М. И., Дерендяев Б. Г. // НТИ. Сер. 2. 1992. №9. С.1–5.
11. Strokov I. I., Lebedev K. S. // J. Chem. Inf. Comput. Sci. 1996. V.36 P. 741–745.
12. www.omnic.com
13. Чмутина К. С., Пиоттух-Пелецкий В. Н. // Современные проблемы органической химии: Тезисы докладов. 17–21 сентября 2001 г. Новосибирск: Издательство СО РАН, 2001. С. 86.

* * * * *

ZAIR, A DATABASE AND SEARCH SYSTEM FOR IR SPECTROSCOPY: ALGORITHMS AND USER INTERFACE FEATURES

I.I.Strokov, K.S.Lebedev, M.I.Podgornaya, B.G.Derendyaev

Original methods of storage and retrieval for IR database containing more than 70000 spectra are proposed. The search algorithm provides fast finding of relevant records by comparison of full curves, peaks, or selected spectral regions.